Census Scientific Advisory Committee Fall 2020 Meeting
Recommendations

To: Ron S. Jarmin
Acting Director
U.S. Census Bureau

From: Allison Plyer
Chair
Census Scientific Advisory Committee

Subject: Remaining Recommendations and Comments to the Census Bureau from the Census Scientific Advisory Committee Fall 2020 Meeting

November 12, 2020

Update on the 2020 Census Operations (continued)

CSAC commends the Census Bureau on their unprecedented efforts to fully enumerate all persons in the United States in the midst of not only a global pandemic and multiple wildfires and hurricanes, but also shifting deadlines, and, in the end, an abruptly shortened Nonresponse Follow-up (NRFU) period. These challenges introduced a need for flexibility and plan changes that are not ideal for the enormous operation that is the decennial census headcount. The Bureau staff adapted to mandated shifting deadlines with professionalism and determination to count everyone in the U.S. once, only once, and in the right place. In light of all that the Bureau had to confront, the completion of data collection for 99.98% of identified residential addresses is impressive.[1]

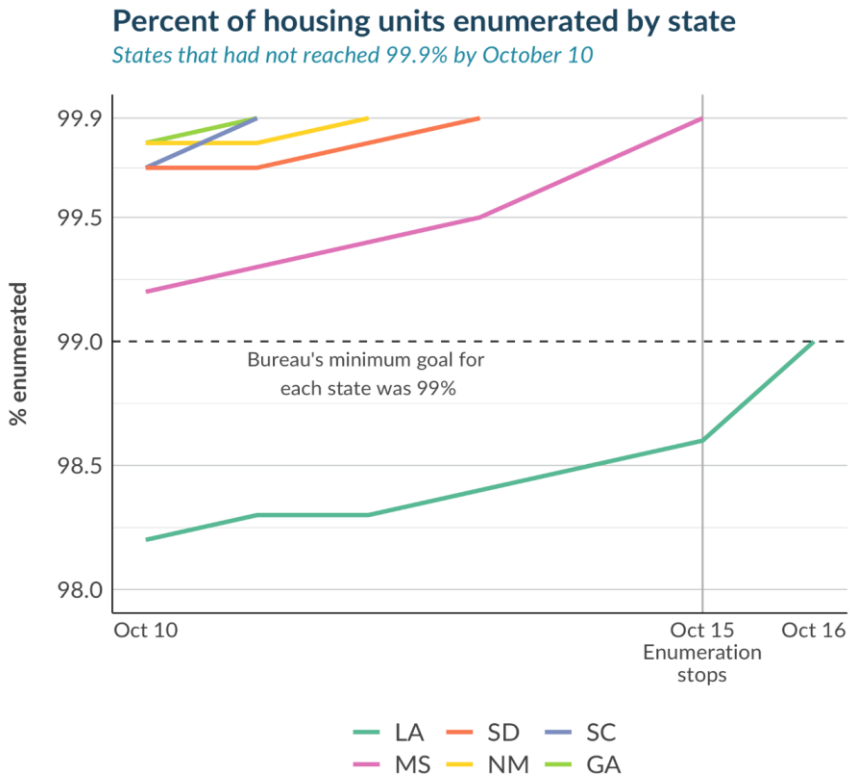
Nonresponse Follow-up

Despite the high completion rate on average nationally, data collection fell short of targets in selected areas around the country, and CSAC remains concerned that:

- Rural communities, given their lower internet access, are particularly at risk of being undercounted by ending NRFU on October 15th. Early assessments indicated that rural communities were lagging urban and suburban communities in self-response -- highlighting the importance of NRFU in ensuring that rural communities are fully enumerated.[2] Rural counties with a high percent of Hispanic/Latinos, Black/African American, and American Indian/Alaskan Native, on average, had lower self-response rates when compared to 2010.

[1] https://www.census.gov/newsroom/blogs/random-samplings.html
[2] https://dailyyonder.com/rural-response-to-census-improves-but-still-lags-national-rate/2020/05/21/

- Wildfires and hurricanes that were active during NRFU undoubtedly affected the Bureau's ability to fully enumerate many communities. Notably, as of October 10, 2020, only one state, that was struck by multiple hurricanes during the NRFU period, had not reached the Bureau's state level completion goal of 99.0%. As of October 15, Louisiana had reached only 98.5% enumerated. Then as of Oct 16 (after the Oct 15 NRFU cut off), the Bureau reported that Louisiana suddenly reached 99% enumeration. At the Bureau's October 21st news briefing, the Associate Director for Decennial Census Programs explained that this sudden jump in Louisiana on October 16th was based on using "high quality administrative data."[3] However, the normal procedure (as explained in the Deputy Director's Nov 5th blog) is to attempt one or more visits, then rely on high quality administrative data, and if no high quality administrative data exists for an address, continue to visit the address or seek out a proxy.[4] Filling in a large number of addresses on October 16th with administrative data suggests those addresses may have had no high quality administrative data, or such data would have been utilized earlier in the process to reduce NRFU attempts. CSAC is concerned these final addresses in Louisiana instead were completed with lower quality administrative data.



[3] https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2020/11/04/shortened-census-led-to-an-incomplete-count-in-some-areas

[4] https://www.census.gov/newsroom/blogs/random-samplings.html

1) Given CSAC's concerns about the completeness and accuracy of NRFU efforts, CSAC recommends that the Bureau publish a set of measures or analyses that would help users determine the fitness of use. Below, CSAC endorses a set of measures proposed by the American Statistical Association's (ASA) 2020 Census Quality Indicators Task Force but understand that the Bureau needs to determine which measures require differential privacy be applied.

   CENSUS BUREAU RESPONSE:

   The Census Bureau accepts this recommendation. The Census Bureau has reviewed and considered the metrics as proposed by the American Statistical Association's Quality Indicator's Task Force. Select metrics will be released on an accelerated schedule at or around the release of apportionment data and at or around the release of redistricting data. Additional metrics will be released as 2020 Census operational assessments and evaluations are completed, as 2020 Census Post-Enumeration Survey results are available, and as additional analysis is completed. All metrics made publicly available will be subject to disclosure review.

2) CSAC recommends that the Census Bureau consult with relevant professional associations about how to balance the need for quality information to determine fitness of use and the interest in preserving most of the privacy budget for the data products.

   CENSUS BUREAU RESPONSE:

   The Census Bureau accepts this recommendation. We are in active collaboration with the American Statistical Association's Task Force and intend to engage a consensus panel from the National Academy of Sciences. Internal and external experts will review confidential operational and preliminary response data and will be afforded some portion of the 2020 Census privacy-loss budget to prepare statistics for public release.

3) In alignment with the ASA's recommendations, CSAC recommends the Bureau release the following field data collection metrics calculated for census tracts, counties, and states:
   - Percent of occupied households enumerated by high quality administrative data
   - Percent of occupied households enumerated by proxy
   - Percent of enumerations that are population count only
   - Percent of occupied households enumerated by administrative data that do not meet the Bureau's standards as high quality
   - Percent of occupied households missing Name or Date of Birth
   - Percent of addresses in the entire NRFU universe that were resolved during the closeout phase
   - Percent of addresses that were unresolved after data collection was concluded

- Percent of NRFU addresses that were enumerated as nonexistent

CENSUS BUREAU RESPONSE:

The Census Bureau accepts this recommendation in part. The Census Bureau is planning to produce, and release select metrics at a national level, by state, for the District of Columbia, and for the Commonwealth of Puerto Rico. All metrics made publicly available will be subject to disclosure review and application of any necessary privacy protections. A determination has not been made regarding the public release of metrics at levels of geography below the state level, or the timing thereof.

Post Data Collection Processing Procedures

CSAC continues to be concerned about the accuracy of the final 2020 Census data based on the time frame for post data collection processing that is now more truncated than envisioned even for the "Replan" compressed 3-month timeline. The current timeline will give the Bureau only 2.5 months to complete these important procedures. CSAC is not aware of any plan for conducting these critical activities in only 2.5 months. Given that large numbers of people changed their normal residential patterns due to pandemic conditions (e.g., college students, snowbirds)[5], adequate de-duplication procedures for college students, retirees, and others require additional time, not less. CSAC understands that the Census Bureau has implemented a seven day/week work schedule in order to meet this shortened timeline.[6] However, after 50 hours/week, worker productivity declines sharply. Moreover, accuracy of the Bureau staff's work is likely to decrease under such sustained pressure.[7] In addition, the quality of group quarters data is reliant on expert review by local state demographers through Count Review Event 2 which was eliminated to accommodate the shortened timeline. Given media reports that some group quarters populations went uncounted,[8] this quality review is critically important to ensure all group quarter populations are correctly enumerated.

4) CSAC continues to recommend that the Census Bureau have the full 6 months it requested in April 2020, to execute its full battery of data checks to reduce the risk of failing to identify key errors and generate final 2020 Census products that are of comparable quality to previous decennial censuses.

[5] https://www.pewresearch.org/fact-tank/2020/07/06/about-a-fifth-of-u-s-adults-moved-due-to-covid-19-or-know-someone-who-did/
[6] https://assets.documentcloud.org/documents/7207428/LUPE-Sept-11-2020-Declaration-of-Albert-Fontenot.pdf
[7] http://ftp.iza.org/dp8129.pdf; https://hbr.org/2006/10/sleep-deficit-the-performance-killer
[8] https://www.sfchronicle.com/news/article/Officials-Vegas-homeless-may-have-gone-uncounted-15692238.php

CENSUS BUREAU RESPONSE:

Since Census Bureau completed field work on October 15, 2020, and the current projected date to provide the numbers to the President is April 30, 2021, this provides 6½ months of post-processing time, more than the amount recommended by CSAC. And just as occurred during data collection, the Census Bureau is continuously checking the quality of the data throughout data processing.

5) In addition, in alignment with the ASA recommendations, CSAC recommends that the Bureau release to outside experts or the public these metrics (based on persons not households) indicating the quality of Post Data Collection Processing efforts tabulated by census tracts, counties and states:

- Percent of records identified as duplicate enumerations across different addresses
- Percent of records that do not contain sufficient information for deduplication
- Percent of records that required status or count imputation
- Percent of records that will require whole person imputation
- Percent of records missing a complete name (first and last)
- Percent of records enumerated from administrative records
- Percent of administrative record enumerations lacking complete names or date of birth

CENSUS BUREAU RESPONSE:

The Census Bureau accepts this recommendation in part. The Census Bureau is planning to produce, and release select metrics at a national level, by state, for the District of Columbia, and for the Commonwealth of Puerto Rico on the same day as apportionment data. All metrics made publicly available will be subject to disclosure review and application of any necessary privacy protections. A determination has not been made regarding the public release of metrics at levels of geography below the state level, or the timing thereof.

6) CSAC recommends that these metrics be released before the release of the first 2020 Census data products, specifically state population counts and apportionment numbers. To accommodate this step toward transparency, the full 6 months for post data collection processing will be necessary.

CENSUS BUREAU RESPONSE:

The Census Bureau accepts this recommendation in part. The Census Bureau is planning to produce, and release select metrics at a national level, by state, for the District of Columbia, and for the Commonwealth of Puerto Rico with apportionment data. All metrics made publicly available will be subject to disclosure review and application of any necessary privacy protections. A determination has not been made regarding the public release of metrics at levels of geography below the state level, or the timing thereof. Lastly, as stated above, Census is taking over 6 months for post-processing.

7) CSAC recommends that the Bureau release the following additional metrics, which are based on the data that will be used to develop the redistricting data:

- Percent of records representing whole person imputations
- Percent of records that required item imputation for Race, Hispanic Origin, Sex, and Age respectively
- Percent of records missing date of birth
- Demographic breakdown of households enumerated by administrative records

CENSUS BUREAU RESPONSE:

The Census Bureau accepts this recommendation in part. The Census Bureau is planning to produce, and release select metrics at a national level, by state, for the District of Columbia, and for the Commonwealth of Puerto Rico on the same day as apportionment data. All metrics made publicly available will be subject to disclosure review and application of any necessary privacy protections. A determination has not been made regarding the public release of metrics at levels of geography below the state level, or the timing thereof.

<u>Partnering with outside researchers</u>

8) CSAC recommends that the Census Bureau provide expedited access to these quality indicators via the Federal Statistical Research Data Centers for analysis that would help the public assess the data quality and fitness of use.

CENSUS BUREAU RESPONSE:

The Census Bureau thanks the CSAC for the recommendation to collaborate with external researchers using the Federal Statistical Research Data Center infrastructure. We concur that this approach will provide additional resources and expertise to analyze the data. We have met with representatives from the ASA and are laying the groundwork to move forward with providing expedited access.

9) CSAC recommends that the Census Bureau contact the American Statistical Association, the Population Association of America, and the National Academy of Sciences (Committee on National Statistics and other divisions) to quickly recommend a small number of researchers to be provided access.

   Providing expedited access to these quality indicators to outside researchers will have the additional benefit of relieving Census Bureau staff of producing additional analyses, during a time when staff are already overwhelmed, working 7 days/week.

   CENSUS BUREAU RESPONSE:

   The Census Bureau thanks the CSAC for the recommendation to collaborate with external researchers using the Federal Statistical Research Data Center infrastructure. We concur that this approach will provide additional resources and expertise to analyze the data. We have met with representatives from the ASA and are laying the groundwork to move forward with providing expedited access.

10) If releasing the geographically detailed quality measures (recommended above) would require the application of differential privacy, CSAC recommends, as an alternative, that outside researchers produce a set of quality measures and analyses that do not require Differential Privacy protection or would require only a very small amount of the privacy budget. These could include:

   - All of these quality metrics at the state level.
   - Correlations between 2010 and 2020 data on the above recommended quality metrics. Such analyses would provide insight into the quality of 2020 data as compared to 2010 data. Concordance correlations would be particularly valuable.
   - Plots in Appendix 2 of the ASA report for 2020 data without identifying tracts. Such plots could cover a large number of variables such as household size, percent African American, percent of records based on Administrative Records, to name just a few.
   - The above recommended quality metrics for demographic groups and/or regions of the country.

These analyses should be produced and made available to the public before and/or with the release of the state population totals and apportionment numbers.

CENSUS BUREAU RESPONSE:

The Census Bureau thanks the CSAC for the recommendation to provide timely release of quality metrics that do not require application of differential privacy or require only a small amount of the privacy budget. We concur with the importance of releasing a series

of quality metrics alongside the apportionment counts. We are in the process of finalizing the list of metrics that will be made available at or shortly after the release of the apportionment counts. Quality metrics will be subject to disclosure avoidance protocols prior to public release.

11) Finally, CSAC recommends the Bureau release its comparison of the 2020 Census counts to the estimates obtained from Demographic Analysis (scheduled for completion in December) as an important, early measure of the 2020 Census data's net undercount or overcount of the national population by race and ethnicity and by age and sex. When releasing the comparison to the estimates from Demographic Analysis (DA), CSAC recommends that the Bureau provide strong cautions regarding the ways in which Demographic Analysis can and cannot be used to assess the accuracy of the 2020 Census results. Specifically, Demographic Analysis cannot be used to determine whether everyone was counted once, only once, and in the right place. Estimates of gross error await the completion of estimates from the Post-Enumeration Survey (PES), which will not be available until 2022.[9] While the 2020 Census results may differ little from the estimates produced by Demographic Analysis, the Census Bureau should demonstrate transparency at the news briefing for Demographic Analysis, by explaining that the final 2020 Census results may be close to the Demographic Analysis numbers, and still many people could be counted in the wrong place, double counted, or not counted at all.

CENSUS BUREAU RESPONSE:

The Census Bureau partially accepts the recommendation. We endeavored to carefully and accurately communicate the utility and limitations of Demographic Analysis during the December 15, 2020 news briefing and Demographic Analysis reports and data products. However, based on the timeline for processing the 2020 Census results, we will be unable to produce and release comparisons of the 2020 Census counts to the estimates obtained from Demographic Analysis prior to the release of the apportionment data. This information will be made available as soon as is feasible, but that will be contingent on when the 2020 Census data files necessary for the analysis are ready for use internally.

12) CSAC recommends that the Bureau release the comparison to Demographic Analysis prior to the release of the first 2020 Census data product, namely state population counts and apportionment numbers.

[9] CSAC notes, however, that Demographic Analysis was more informative than the PES in documenting the net undercount of young children in the 2010 Census (U.S. Census Bureau, "The Undercount of Young Children," February 2014).

CENSUS BUREAU RESPONSE:

The Census Bureau partially accepts the recommendation. The official 2020 Demographic Analysis estimates are limited to the national level. It is our intention to release a report that contains a comparison between the 2020 Demographic Analysis estimate for the nation and the apportionment total as close to the day the apportionment data are released as possible. This report will also feature the first calculation of net coverage error.

13) Echoing the recommendations from the Partnership for America's Children, CSAC recommends that the Bureau produce a report based on the Demographic Analysis (which was not done in 2010).

CENSUS BUREAU RESPONSE:

The Census Bureau accepts the recommendation and plans to produce a report in 2021 exploring the results of the 2020 Demographic Analysis and estimates of net coverage error. The timing of this report will be contingent on when the 2020 Census data files necessary for the analysis become available internally.

14) CSAC also recommends that the Bureau release comparisons of 2020 Census data to Census Bureau population estimates available for states, cities, and towns to identify large discrepancies at smaller geographies as an indicator of possible over or undercounts. CSAC understands that comparisons below the state level will be affected by the application of differential privacy, but the results of such comparisons remain an important measure of 2020 Census data quality. The Bureau could enlist the assistance of outside researchers with these analyses, given that sufficient internal resources may not be available.

CENSUS BUREAU RESPONSE:

The Census Bureau partially accepts this recommendation to use the Vintage 2020 population estimates to evaluate the quality of the 2020 Census. Importantly, the decennial census is more commonly used to evaluate the quality of the estimates, which are based on the previous census and designed to capture population change across the decade. Thus, while it is our intent to make comparisons between subnational population estimates and the 2020 Census counts, it will need to be strongly emphasized that any differences observed may be because of the estimates methodology, and are not necessarily reflective of the quality of the decennial census. These analyses will be conducted as part of our Estimates Evaluation project, the details of which are still being finalized.

15) CSAC recommends that the Bureau release the comparisons to state population estimates on the same day as the release of the first 2020 Census data product, namely state population counts and apportionment numbers.

CENSUS BUREAU RESPONSE:

The Census Bureau accepts the recommendation. It is our intention to release a report that contains a comparison between the Vintage 2020 population estimates for states and the apportionment totals on the same day as the release of the state population counts and apportionment numbers, or as close to that day as possible.

16) Consistent with Census Bureau practices since 1950, CSAC recommends that the Census Bureau announce the total populations of each state in a press briefing at the National Press Club, while also announcing the apportionment numbers, and on the same day transmit these numbers to the President.[10]

These efforts toward transparency will optimize the American people's confidence in the 2020 Census data, in all Census Bureau surveys, and in the Census Bureau itself. Notably, the share of Americans who doubt that the 2020 Census results would be accurate grew from 24% in early March 2020 to 31% by late July.[11] Radical transparency is essential for restoring Americans' trust in the Census and securing the Bureau's reputation. Transparency increases credibility and loyalty among customers. While institutions may go to great lengths to accentuate the benefits of their products and downplay risks, being transparent, "warts and all," about challenges and limitations has the desired effect of building credibility and loyalty.[12] In short, transparency is a mission-critical part of the Census Bureau's operations.

CENSUS BUREAU RESPONSE:

The Census Bureau thanks CSAC for this recommendation and will, to the extent feasible, have an apportionment release where the content and format are consistent with this recommendation. The Census Bureau concurs with the importance of transparency for confidence in the 2020 Census data, and the Census Bureau plans to release the apportionment data on the same day they are delivered to the President. The apportionment press briefing may be virtual, rather than in person due to COVID-19. We will let the public know as soon as final decisions are made.

17) CSAC recommends that Bureau leadership and communication staff avoid the temptation to gloss over the "warts" of the 2020 Census process and data, and instead

[10] https://apportionment.info/
[11] https://www.pewresearch.org/fact-tank/2020/07/28/four-in-ten-who-havent-yet-filled-out-u-s-census-say-they-wouldnt-answer-the-door-for-a-census-worker/
[12] https://hbswk.hbs.edu/item/the-upside-of-highlighting-a-products-downsides

focus on providing an unprecedented level of transparency for the 2020 Census -- which was conducted under unprecedented circumstances.

CSAC commends the Bureau for the Deputy Director's blog of November 5, 2020 which provided transparency around national metrics as well as questions yet unanswered.[13] CSAC looks forward to the Bureau's release of these and other important measures and analyses of accuracy and quality of the 2020 Census processes and data.

CENSUS BUREAU RESPONSE:

The Census Bureau concurs in the need to provide an unprecedented level of transparency surrounding the quality of the 2020 Census.

Post-Enumeration Survey Methodology

CSAC would like to thank the Census Bureau for the presentation on the Post-Enumeration Survey (PES), its history, and its basic design. CSAC appreciates the importance of the PES in evaluating the quality of the 2020 Census by measuring net coverage and estimating the gross components of coverage: correct enumerations, omissions, erroneous enumerations, and whole person imputations. The circumstances for the 2020 Census are unusually challenging. Added to the expected challenge of declining survey response rates[14], the 2020 Census has faced additional challenges from a pandemic, natural disasters, late changes to processes, and accelerated timetables. This unprecedented combination of challenges makes the importance of the PES even greater than for other decennial censuses. At the same time, the challenges affecting the 2020 Census are likely to also affect the PES.

1) CSAC recommends that the Census Bureau provide maximum possible transparency on process indicators for the PES. Such transparency will increase trust in the final results and may enable the Census Bureau to obtain useful feedback and help throughout the implementation of the PES.

CENSUS BUREAU RESPONSE:

The Census Bureau accepts this recommendation. We are committed to being transparent and look forward to presenting quality indicators and PES results.

[13] https://www.census.gov/newsroom/blogs/random-samplings.html
[14] Czajka, J. L., & Beyler, A. (2016). Declining response rates in federal surveys: trends and implications (background paper). Mathematica Policy Research.

2) CSAC also recommends that the Census Bureau consider how to separate what is learned from the PES regarding temporary challenges (e.g., the pandemic), versus ongoing challenges (e.g., declining response rates).

Because the PES data collection occurs after the NRFU data collection, which was delayed by three months due to the pandemic, CSAC is concerned about the potential impact on the PES data collection and how this might affect the 2020 PES coverage evaluation. CSAC is concerned, in particular, about two possibilities: 1) The PES may encounter sharply higher than expected nonresponse. 2) Because of the longer than usual time lag between census day and the PES data collection, the quality of the PES data obtained from all modes may be lower than the quality of the corresponding Census data. Either of these possibilities would seem to have the potential to result in the PES overestimating incorrect enumerations, whether or not this affects the PES estimate of net Census error.

In view of the challenges faced by the PES, CSAC requests that the Census Bureau provide a detailed update at its spring 2021 meeting so that CSAC members can review the Bureau's progress and have an opportunity to offer suggestions to address outstanding methodological issues, including development of a suitable correction for correlation bias among children. In addition, while the Census Bureau has published an operational plan for the 2020 PES, a detailed description of the PES methodology is wanting. CSAC understands that the methodology will be similar to what was employed for the 2010 Census but given that the 2010 implementation was innovative in a number of respects, CSAC wonders if the 2020 PES will include any innovations or refinements based on what was learned from the 2010 application as well as improvements in matching technology and expanded computing capacity.

CENSUS BUREAU RESPONSE:

The Census Bureau accepts the recommendation to consider how to separate what is learned from the PES regarding temporary challenges versus ongoing challenges. In April, a team was formed to document changes the PES made in response to the COVID-19 pandemic and the potential impact of COVID-19 on the PES. Four of the five PES field data collection operations were delayed because of COVID-19. As the CSAC has noted, this could have an impact on the quality of the PES estimates. The Census Bureau has taken measures to mitigate many potential errors, especially related to response and recall error.

The Census Bureau would be glad to present an update on the PES design and operations at the next CSAC meeting.

3) CSAC recommends that the Census Bureau produce documentation on the 2020 PES methodology prior to the spring CSAC meeting so that CSAC members may have an opportunity to review the methodology and offer informed commentary.

   More immediately, CSAC asks to be apprised of the schedule for completing the PES (for conducting the survey, performing the matching and estimation, and releasing the final results) to the best degree it is known at the time of the Census Bureau's response. CSAC also asks to be informed if the Census Bureau is considering any methodological adjustments to the PES given the unprecedented context. For example, is the Bureau making any changes to how respondents are contacted or introducing any special procedures to deal with the college students who are included in the PES universe (those who were not living in dormitories).

   CENSUS BUREAU RESPONSE:

   The Census Bureau accepts the recommendation to provide documentation about the 2020 PES methodology. The design of the 2020 PES memo documents the intended design of the PES, prior to the COVID-19 pandemic. Two additional memos from the PES quality team that document a number of the changes and impacts resulting from the COVID-19 pandemic are in review and can be released upon final review.

   The following table shows new dates for the PES field operations. Note that the Independent Listing operation schedule was not impacted by the COVID-19 pandemic. The Person Interview data collection was extended to give more time to increase responses, especially in areas that were closed due to COVID-19.

Operation	Original Dates		New Dates	
	Start	Finish	Start	Finish
Independent Listing	1/16/2020	3/13/2020	N/A	N/A
Initial Housing Unit Followup	5/6/2020	6/19/2020	7/30/2020	9/21/2020
Person Interview	6/17/2020	9/18/2020	9/23/2020	3/20/2020
Person Followup	2/3/2021	3/26/2021	6/14/2021	8/12/2021
Final Housing Unit Followup	5/19/2021	6/18/2021	10/25/2021	11/30/2021

The Census Bureau currently plans to disseminate the first set of census coverage estimates for people late in 2021 or early in 2022, and a second set of census coverage estimates for people and housing units in the spring of 2022.

Contact mode has changed in response to COVID-19. The Census Bureau would be glad to provide details about these changes, many of which are in the two memos from the PES Quality Team.

4) CSAC also looks forward to the Census Bureau's release of the PES results in 2022 and its use in the analysis of the 2020 Census count's quality including undercounts and overcounts. Consistent with the recent recommendations from ASA, CSAC recommends that the Bureau consider how it might use the results of its coverage measurement program to take actions to make corrections, if needed, to the 2020 Census based data products to optimize the accuracy of these products before another full enumeration of the nation takes place. Such corrective action could improve funding allocations to areas undercounted by the 2020 Census enumeration.

CENSUS BUREAU RESPONSE:

The Census Bureau does not accept this recommendation. The objective of the PES is to evaluate the census. The PES was not designed to correct any census products. Numerous logistical, operational, and technical challenges make this infeasible.

Pulse Surveys

CSAC commends the Bureau for the incredible efforts in devising the program of Pulse surveys. Large scale (including regional) disasters create an enormous break in the status quo followed by months and often years of flux. By their very nature, disasters create confusion and rumors start to circulate post-disaster. After a disaster there is escalated demand for timely data to address rumors, assess the current status, and re-assess over time—all to inform and catalyze critical community, planning, and investment decisions that support recovery. Indeed, post-disaster data can actually catalyze investments by reducing uncertainty, and as such is a very important contributor to recovery. But just as demand for data escalates after a disaster, leaders complain of a vacuum of data after a disaster. This is because most data are insufficiently timely to be relevant after a disaster.

The Bureau's work to address pressing questions in a timely manner soon after the COVID crisis began is an extremely important example of the kind of data collection that is essential following each disaster. Updating this data frequently is critical as well. As conditions change, the questions may change, and the frequency with which data is collected can decrease over time. The Bureau is learning these lessons about the dynamics of post-disaster data demand through their own direct experience.

CSAC commends the Bureau for considering institutionalizing rapid response surveys for future national or regional emergencies. Such efforts can go a long way to accelerating disaster recovery. While nearly all large-scale disasters are similar in generating demand for data, disasters often differ in the types of data needed.

To identify the highest demand data, The Data Center of Southeast Louisiana uses a version of Design Thinking that the Bureau could consider implementing. It entails scanning local media for descriptions of lived experiences (e.g., small businesses closing, people unable to return to their homes, workers becoming disconnected from employers) combined with demographic and economic expertise to develop hypotheses about whether such experiences are likely to be common or widespread. The Data Center prioritizes gathering data on those experiences or issues that specifically can inform recovery, planning and investment decisions.

Gathering local knowledge about decisions that are being made with no data or bad data, would be another way to identify and prioritize data collection post-disaster. The response period is short-lived and the Bureau's data collection efforts are likely to be more impactful if focused on informing recovery efforts rather than immediate response. The Bureau could reach out to local planning departments of disaster-affected municipalities to identify the decisions they are struggling to make because of lack of data.

To create a question bank in advance, the Bureau could assemble a number of leaders who have hard-earned, on-the-ground expertise in disaster recovery following common disasters such as wildfires and hurricanes to identify some of their most common data needs. These should include the data needs for understanding the implications for children and youth as well as adults. The National Low-Income Housing Coalition currently convenes a group of such experts on a regular basis. Resilient Cities Catalyst is a Rockefeller Foundation initiated organization that supports recovery and resilience across many cities and could assemble a group of experts to identify high impact data needs. School districts, whose operations were particularly hard hit by COVID, could also be consulted. Many cities have Chief Resilience Officers who could provide good inputs. The Bureau could also get input from long-term resilience committees and longer-term recovery committees that have now been established by many states.

1) CSAC recommends the Bureau consider a TOP (The Opportunity Project) Sprint that engages diverse stakeholders recovering from disasters. Because individuals from the most vulnerable communities are almost always the most deeply impacted by disasters, TOP could specifically engage representatives from such communities -- particularly those with access and functional needs, lower income communities, households with language barriers, renters, etc.

CENSUS BUREAU RESPONSE:

The Census Bureau accepts this recommendation and is already engaged with The Opportunity Project and will be presenting both the Small Business Pulse and Household Pulse surveys as project areas for the TOP sessions.

2) Because it is based on the Master Address File (MAF), the Household Pulse Survey (HPS) can be linked to other Census data like the American Community Survey. When resources permit, CSAC recommends creating a (restricted) version of the HPS matched with other data at a fine geographic level (e.g., Census tract) for each respondent. This would allow researchers to understand which characteristics of communities helped predict relative success or failure in confronting the pandemic.

CENSUS BUREAU RESPONSE:

The Census Bureau accepts this recommendation for the Household Pulse Survey. Researchers and data users can currently access restricted HPS data via an approved project through the Federal Statistical Research Data Centers. Even without linking the HPS data with other Census data, a user could match to Census tracts or other finer geographic level based on MAFID. However, HPS sampling is not designed to be representative at the substate level. There is also uncertainty in the address because the sampling MAFID is not accurate in all cases for the collected data. There are people who have moved or whose email or phone number was associated with an incorrect address. While we attempt to assign them to the correct state for weighting, we do not assign lower-level geography, but a user in the Federal Statistical Research Data Centers could attempt to geocode based on the respondent provided address.

3) The pandemic (and other shocks) are being experienced differently in and out of metropolitan areas. Reliance on public transportation and the quality of medical facilities, for example, differ in metropolitan and non-metropolitan areas. At present, the Household Pulse Survey is representative for states and for the 15 largest Metropolitan Statistical Areas (MSAs). If resources permit, CSAC recommends making future Pulse Surveys representative for MSA status by state (or at least by Census region).

CENSUS BUREAU RESPONSE:

The Census Bureau will take this recommendation under advisement. We will discuss the sample requirements versus the cost of collecting the data for these geographies.

4) The weekly/biweekly frequency of the Pulse data is touted as an advantage. With the two phases now completed, CSAC recommends formal tests for the incremental value of having the data at a weekly/biweekly rather than monthly frequency. Indicators from

the Pulse Surveys presented in the one-way briefing in August did not, in general, reveal sharp week-to-week differences.

CENSUS BUREAU RESPONSE:

It is hard to know a priori when a shock will result in economic activity changing at a weekly, bi-weekly, or monthly frequency. Early in the pandemic, federal, state, and local policies were moving quickly so it made sense to have a weekly collection. The problem is that while we are in the moment, we cannot accurately forecast the likelihood of policy action. In addition, we are not able to forecast a change in the underlying cause of policy actions: the effect of the Coronavirus pandemic on the economy. We cannot predict changes in the severity of the pandemic (e.g., will it worsen in flu season?) nor future developments that will alleviate the pandemic (e.g., vaccines or treatments). In a period of such high uncertainty, the impossibility of forecasting these inflection points underscore the benefits of having a weekly survey. For these reasons, the Census Bureau will proceed with a weekly collection for the Small Business Pulse Survey.

The Household Pulse Survey followed similar reasoning for the initial weekly design but has shifted to a biweekly design with the expansion of content in Phase 2. The primary benefit is the reduction in respondent burden associated with fewer cycles and the conservation of sample for ongoing operations. The HPS is expected to retain its biweekly collection cycles.

5) CSAC further recommends that the Pulse Surveys include some identical questions to other existing administrative data sets and surveys available on the national, state, and local levels, as a means of cross validating the Pulse Surveys. At the national or state level, such administrative data sets and economic surveys might include Initial UI Claims (weekly), Current Population Survey (monthly), Current Employment Statistics (monthly), the National Survey of Children's Health, and ISM Reports on Business (monthly). The need for cross-validation is particularly important given the understandably low response rates on the Pulse Surveys. Once validated, the Pulse Surveys can be used to augment the key elements of those existing surveys and administrative data sets on a timelier basis during times of national or local emergency. Further, the wording of questions in the Pulse Survey could provide a template for state and local jurisdictions to implement their own surveys as local conditions warrant.

CENSUS BUREAU RESPONSE:

The Census Bureau will take this recommendation under advisement. This is a longer-term objective to compare survey results to administrative files. For example, for the Small Business Pulse Survey, we are evaluating IRS tax data on business deaths and aim to have a product on this topic in 2021. However, there are administrative files we do

not currently have, like financial data such as loans and payment defaults, but we are working with the Small Business Administration to acquire this information and make those linkages.

The recommendation will be taken into consideration for future iterations of the Household Pulse Survey. Survey content is currently developed in partnership with our partnering agencies. Most of these agencies are sponsors of existing Census Bureau surveys and many of the questions are pulled, either exactly or slightly modified, from these existing surveys, allowing cross validation of results. For example, the National Center for Health Statistics included the question, "Over the last 7 days, how often have you been bothered by the following problems ... Feeling nervous, anxious, or on edge? Would you say not at all, several days, more than half the days, or nearly every day?" This question was modified from a similar question in the National Health Interview Survey, which asks, "How often do you feel worried, nervous or anxious? Would you say daily, weekly, monthly, a few times a year, or never?" Formal plans to integrate the Household Pulse Survey with other data systems have not been discussed, partly as a result of the uncertain duration for the continuation of the data collections.

6) Both Pulse surveys emphasize questions whose answers would be hard to obtain without household interviews. This is the comparative advantage of the Pulse surveys relative to other datasets that have been used to study COVID's effects. Since so much of any given household's experience depends on how the whole community responds to COVID-19, CSAC recommends including questions in the Household Pulse survey on the household's perception of the community's adherence to wearing masks, social distancing, quarantining, etc. Questions on internet access could also be added, beyond its role in primary and secondary education, given its importance during the lockdown periods. Additionally, since so much of the federal government's assistance to small businesses was in the form of potentially forgivable loans administered through banks, the health of the local banking sector is important to understand.

CENSUS BUREAU RESPONSE:

The Census Bureau will take this recommendation under advisement. In regard to the last sentence: "Additionally, since so much of the federal government's assistance to small businesses was in the form of potentially forgivable loans administered through banks, the health of the local banking sector is important to understand," we agree this is something that would be valuable to collect, but it needs to be weighed against the importance of other concepts collected on the Small Business Pulse Survey. In order to introduce new content, we must justify the need with the Office of Management and Budget. Additionally, cognitive testing is required to ensure the new questions achieve the intended outputs. We have limited resources and also need to take into account the burden on businesses to respond. We have worked with the Federal Reserve Board and

the Small Business Administration to collect information that is useful to their stakeholders and the businesses and policymakers that they serve.

Similarly, the Household Pulse Survey has the same constraints of OMB approval, testing needs, and respondent burden.

7) CSAC recommends that future versions of the Small Business Pulse Survey include questions about the performance of the banking sector in administering these programs.

The Pulse Surveys reflect the Census Bureau's comparative advantage in data collection – direct surveys of representative samples. There have been many other data collection efforts, relying more often on passively collected data that are not necessarily representative of the total population. Some examples include transaction-level data from financial intermediaries, social network data aggregated by locality, and smartphone location data.

CENSUS BUREAU RESPONSE:

See response to question 6.

8) CSAC recommends that when time and resources permit, the Census Bureau should compare the conclusions drawn with each type of data and integrate them into a comprehensive report on the pandemic.

CENSUS BUREAU RESPONSE:

The Census Bureau will take this recommendation under advisement. As time and resources permit, we will detail our collective response to the pandemic.

9) Lastly, CSAC commends the Bureau for offering the Pulse survey in Spanish. However, the instructions for Spanish speakers are in English, which may be an obstacle for Spanish language speakers.

CSAC recommends that the Bureau make the instructions as well as the Pulse survey itself available in Spanish.

CENSUS BUREAU RESPONSE:

The Census Bureau accepts this recommendation. For the Household Pulse Survey, the instrument and instructions are currently available in both English and Spanish. Approximately 1% of the questionnaires are completed in Spanish.

Construction Modernization

CSAC would like to thank Stephanie Studds and her team for such impressive work. The use of innovative techniques, such as the change detection from satellite images, allows the team to modernize the collection and analysis of construction data.

The goal of the Construction Modernization project is to reengineer the measurement approach to the traditional construction surveys by utilizing alternative data sources, developing modeling techniques, and evaluating the use of satellite technology.

The Economic Indicators Division of the Census Bureau has been tasked with the modernization effort and has made significant progress modernizing the workflow. The Working Group's focus is to work with the Census Bureau to provide input to assist with the following:

- Developing methodologies to maximize data consistency, accuracy and geographic coverage and granularity
- Reducing the need for field collection and/or burden to respondents
- Implementing a methodology for real-time data ingestion and updates
- Implementing a methodology that remains cost neutral across the construction programs
- Defining key milestone markers, such as what indicators show completion of construction
- External communication related to the impact of the program
- Communication of findings and reports

The Construction Modernization Working group currently consists of four CSAC members, all of whom will complete their terms on the CSAC in Spring of 2021. There is a need for additional CSAC members to participate in the working group to provide continuity and a means to transition the effort.

The Construction Modernization project's activities make it evident that partnerships with the private sector, particularly industries of interest such as insurance and housing, could add significant value to the expected products. Other Census Bureau data products could also benefit from these activities. CSAC encourages internal communication and coordination across the Bureau. For example, the tempo of the urban/rural classification of the census blocks could be enhanced by the utilization of these data.

1) CSAC recommends that the Census Bureau continue the Construction Modernization – Re-engineering Initiative and include a follow-up presentation of results at the next CSAC meeting.

CENSUS BUREAU RESPONSE:

The Census Bureau appreciates your interest and feedback on the construction modernization and re-engineering initiative. We will provide a follow-up presentation at the spring meetings.

2020 Census CVAP (Citizen Voting Age Population) Special Tabulation

CSAC commends the Bureau for the extensive effort that went into securing access to administrative data from various sources, matching those records to prior census and survey records, assigning citizenship when administrative records contain sufficient information to do so, and formulating and testing various methodologies for imputing citizenship to those census or survey records that either did not link or otherwise could not be assigned citizenship status from administrative records. CSAC notes that the resolution (with high certainty) of a large proportion of cases via linkage and the application of business rules is by itself a substantial advance over the American Community Survey (ACS) self-reported citizenship.

Business rules are procedures designed to determine assignment of citizenship status based on varied administrative data sources. These can only be applied to records that can be matched to administrative records. CSAC agrees with the Bureau that the remaining 9% of cases that cannot be resolved via business rules are fundamentally different from the business rules cases, even conditionally on covariates. Given this nonignorable missingness, CSAC agrees that imputation methods trained on business rules cases (hot-deck and logistic regression) are inappropriate for predicting non-business rules (NBR) cases, and the best option among those considered is the logistic regression trained on ACS NBR cases. This option essentially predicts what ACS self-reported citizenship would have been for Census NBR cases, assuming Census NBR respondents behave like ACS NBR respondents. The Bureau could explain to a lay audience that this model was selected because it relies on actual administrative records, and when administrative records could not be accurately matched to a census respondent, it models that respondent's likely response to the ACS question about citizenship status. CSAC notes that the choice of model is critical and each yield different results for different demographic subgroups.[15]

Importantly, there is no accurate external data set (that is, no "ground truth") for the self-reported citizenship in the ACS NBR cases, so the degree of misreporting in these cases is unknown, and the extrapolation to Census NBR cases contains further unknowns. The lack of ground truth is not a unique disadvantage because past CVAP tabulations were also based on ACS self-reported citizenship, but lack of ground truth means there are no testing data on which to base objective measures of uncertainty.

[15] https://www2.census.gov/ces/wp/2020/CES-WP-20-33.pdf

1) CSAC recommends that the Bureau critically assess the fitness-for-use of the CVAP data it produces given that the 9% of the population that has the imputed citizenship attribute comes with large error bars. One measure of the (un)fitness would be the discrepancy between the numbers if different methods are used for imputation. It could be useful, e.g., if the published numbers made clear that the specific derived numbers could vary in non-trivial ways if we chose one way or the other from reasonable-looking models.

CENSUS BUREAU RESPONSE:

In accord with EO 13986, the Census Bureau no longer plans to produce a CVAP using administrative records data, so this recommendation is moot.

With the exception of the hot deck method, which the technical report does not recommend using, the methods we have tested do not conduct imputation (assigning 0's and 1's to the citizenship variable) for NBR cases. Rather, they produce citizenship probabilities between 0 and 1.

2) CSAC also recommends that the Bureau continue to explore potential biases in self-reported citizenship for NBR cases within the ACS, as a means of studying the quality of the CVAP tabulations. CSAC encourages the Bureau to engage with credible, independent research organizations or agencies for further study on the quality of ACS self-reported citizenship for NBR cases, e.g., draw a subsample of ACS NBR cases and follow up with interviews.

The Latent Class model is still under development and has not been assessed by CSAC. From the Bureau's presentation, the Latent Class model does not rely explicitly on business rules, but it does rely implicitly on those rules. Cases that cannot be resolved via business rules will continue to be problematic in the Latent Class model due to fundamental limitations of the data, and the model should reflect that difficulty with larger prediction uncertainties for NBR cases.

CENSUS BUREAU RESPONSE:

The Census Bureau accepts the recommendation. We plan to do further study of the quality of ACS self-reported citizenship. We will investigate the possibility of conducting follow-up interviews for ACS NBR cases. Note, however, that this would not be easy to do successfully, given the sensitivity of the citizenship question.

We can study ACS self-reported citizenship quality in other ways as well. To assess the quality for those with sufficient personally identifiable information (PII), we can compare the ACS citizenship responses to administrative records for cases where newly

accessed administrative citizenship data move the cases from NBR to BR cases. We can study citizenship response quality of persons with insufficient PII by comparing their ACS citizenship responses to administrative records linked to the ACS person record by address. Though it may not be possible to distinguish discrepancies due to incorrect linkage vs. ACS misreporting, we should at least be able to detect the direction of any ACS misreporting bias.

CENSUS BUREAU RESPONSE:

The LC model makes no formal distinction between cases that are business-rule resolved and unresolved. Rather, it computes a posterior probability of citizenship for each person, conditionally given that person's characteristics and citizenship measures. Our analyses have shown that, for the vast majority of cases deemed resolved by the rules-based methods, these posterior probabilities lie very close to zero or one, so the LC model effectively treats them as resolved. Cases deemed unresolved by the rules-based methods tend to have posterior probabilities that are less extreme, conveying greater uncertainty. Posterior probabilities far from zero or one are not inherently problematic if the model's assumptions are true, but (as noted by CSAC in Point 3 below) those assumptions should be carefully evaluated.

3) CSAC recommends careful analysis of future results from the Latent Class model to ensure that prediction uncertainties are realistic and are not driven by untenable modeling assumptions. Notably, the Latent Class model will be the most difficult to describe to a lay audience. Understanding the model is very important. The assumptions behind this model need to be clearly stated in a publicly accessible document.

CENSUS BUREAU RESPONSE:

The Census Bureau accepts this recommendation. We are currently exploring ways to check the LC model's assumptions. The assumption of local independence can be relaxed in certain ways (for example, by allowing residual associations among pairs of items within classes) and although this needs further study, we do not believe that these associations would appreciably change the results. More importantly, we believe it is crucial to examine covariates that may moderate the relationships between certain items and class membership. An enhanced version of the LC measurement model that incorporates moderator variables is under development. Describing the LC model in an accessible way to a non-technical audience is challenging but essential. Analogies to situations that are becoming more familiar to the public—for example, the probability that an individual is infected with COVID-19 given positive or negative results from one or more imperfect diagnostic tests—may be helpful in this regard.

4) The Bureau has stated that block-level CVAP tabulations will be protected as part of the overall Census differential privacy procedure, with a portion of the privacy loss budget to be determined. CSAC recommends that the privacy loss budget be set at a small value (high protection, low accuracy) with the recognition that no use case requires block-level CVAP data; CVAP data must be aggregated into geographies that cannot be specified in advance. Since any allocation of privacy loss budget to CVAP necessarily reduces accuracy for other Census data products, CSAC supports this allocation plan for the CVAP privacy loss budget.

CENSUS BUREAU RESPONSE:

The Census Bureau notes the recommendation. The Data Stewardship Executive Policy (DSEP) will be advised of this recommendation when it makes final determination of the privacy-loss budget for the 2020 Census data products.

5) One remaining concern with the aggregation of block-level, differentially-private estimates to novel geographies is that positive biases induced by post-processing in the TopDown Algorithm (due to the nonnegativity constraint) could add up non-trivially in the new geography. CSAC recommends that the Bureau study this potential for additive biases in the CVAP block-level data.

CENSUS BUREAU RESPONSE:

In accord with EO 13986, the Census Bureau no longer plans to produce a CVAP at the block level using administrative records data, so this recommendation is moot.

6) Because it is important to understand the impact of imputation, CSAC recommends that the Census Bureau provide a summary breakdown of its citizenship estimates into four subgroups:
   1. Business Rules (BR) citizen
   2. BR non-citizen
   3. Imputed citizen
   4. Imputed non-citizen

   and present these results for each of the four modeling methods and, if available, for both the 2010 CEF and the 2018 ACS. The CVAP PowerPoint reports some of these subgroups based on the 2010 CEF and the rest based on the 2018 ACS. The technical paper doesn't provide this breakdown either.

CENSUS BUREAU RESPONSE:

In accord with EO 13986, the Census Bureau no longer plans to produce a CVAP using administrative records data, so this recommendation is moot.

7) CSAC recommends that the Bureau produce ACS-based CVAP estimates at least for 2020 in parallel with the 2020 CVAP estimates so that researchers have a bridge in the data that will allow for them to do comparisons overtime. In addition, the Bureau should point to the break in time series and change in methodology with appropriate cautions about drawing conclusions between the distinct estimates.

CENSUS BUREAU RESPONSE:

In accord with EO 13986, the Census Bureau no longer plans to produce a CVAP using administrative records data, so this recommendation is moot. Only the ACS-based CVAP product will be produced.

8) CSAC welcomes the release of the technical paper. CSAC recommends that the Bureau publish additional technical papers as well as journal articles to describe their methodology and the basis for their choice. CSAC recognizes that documenting methodologies and sharing them externally require substantial time investment and we applaud the Census Bureau on their efforts to share their work with external stakeholders.

CSAC acknowledges that conveying uncertainty reflecting prediction error, which is not due to sampling error, is important. CSAC encourages the Bureau to continue to grapple with this important question and to share their conclusions at a future CSAC meeting.

CENSUS BUREAU RESPONSE:

The Census Bureau accepts the recommendation; however, in accord with EO 13986, the Census Bureau no longer plans to produce a CVAP using administrative records data, so this recommendation is interpreted as encouraging the Census Bureau to continue to expose its research to the scientific community in appropriate technical reports and journal articles. We plan to conduct research on the uncertainty of the model-based administrative record estimates and publicly release the findings. The estimates of uncertainty will include the contribution from the differentially private disclosure avoidance system. The LC model is likely to play a crucial role in assessing prediction error, because its Bayesian formulation provides a coherent framework for computing measures of uncertainty in modeled statistics at any level of aggregation. We will prioritize submitting an article on the technical details of the LC model to a quality peer-reviewed journal and a second article on the software developed for this project to a peer-reviewed journal that specializes in computation.

Public Comments

CSAC appreciates the Census Bureau enabling public engagement and recommends that the Census Bureau respond in writing to the two written public comments.

CENSUS BUREAU RESPONSE:

Thank you for your recommendations.